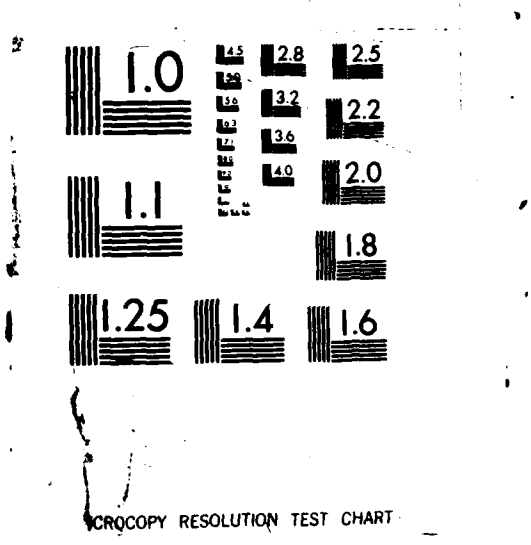
AD-A174 663 KNOWLEDGE DELIVERY RESEARCH(U) UNIVERSITY OF SOUTHERN 1/1
CALIFORNIA MARINA DEL REY INFORMATION SCIENCES INST
W C MANN OCT 86 ISI/SR-86-178 F49620-84-C-0100

UNCLASSIFIED F/G 9/2 NL


1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART

ISI Special Report
ISI/SR-86-178
October 1986

(10)

University of Southern California

William C. Mann

Final Report: Knowledge Delivery Research

INFORMATION SCIENCES INSTITUTE

213/822-1511
4676 Admiralty Way/Marina del Rey/California 90292-6695

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS AD-A174 663
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT This document is approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ISI/SR-86-178	5. MONITORING ORGANIZATION REPORT NUMBER(S) ----------
6a. NAME OF PERFORMING ORGANIZATION USC/Information Sciences Institute 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION ----------
6c. ADDRESS (City, State, and ZIP Code) 4676 Admiralty Way Marina del Rey, CA 90292	7b. ADDRESS (City, State, and ZIP Code) ----------
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Advanced Research Projects Agency 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER FQ86718401007
8c. ADDRESS (City, State, and ZIP Code) 1400 Wilson Boulevard Arlington, VA 22209	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. ---------- PROJECT NO. ---------- TASK NO. ---------- WORK UNIT ACCESSION NO. ----------
11. TITLE (Include Security Classification) Final Report: Knowledge Delivery Research (Unclassified)	
12. PERSONAL AUTHOR(S) Mann, William C.	
13a. TYPE OF REPORT Research Report 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) 1986, October 15. PAGE COUNT 15
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
09	02	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
artificial intelligence, coherence, computational linguistics, discourse, discourse analysis, generation, grammar,
(over)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The goal of knowledge delivery research is to create a technology of authorship by computer. Existing technology is all in the laboratory stage, and is limited to very small, rigidly constrained texts. This research project has focused on two kinds of developments:

1) expanding the notation and practices of knowledge representation so that a wider range of knowledge can be rendered in natural language, and

2) creating a theory of text structure that is suitable as a basis for writing programs that design text.

Keywords:

This is the final research report for AFOSR contract FQ8671-84-01007, an ongoing research program at USC Information Sciences Institute. The research is not complete, and is being continued under contract F49620 87 C 0005. This report represents the research accomplishments in the interval of August 15, 1984 to August 14, 1986

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☑ UNCLASSIFIED/UNLIMITED ☑ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Sheila Coyazo Victor Brown	22b. TELEPHONE (Include Area Code) 213-822-1511 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

(continued from 18.)

knowledge delivery, linguistics, natural language processing, rhetorical structure theory, systemic linguistics, text generation, text structure, text planning, writing

ISI Special Report
ISI/SR-86-178
October 1986

William C. Mann

Final Report: Knowledge Delivery Research

INFORMATION SCIENCES INSTITUTE

213/822-1511

4676 Admiralty Way/Marina del Rey/California 90292-6695

This research was supported by the Air Force Office of Scientific Research Contract No. FQ8671 84 01007. The views and conclusions in this report are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either explicit or implied, of the Air Force Office of Scientific Research of the U.S. Government.

Final Report
Knowledge Delivery Research

William C. Mann
USC/ISI

October, 1986

This research was supported by the Air Force Office of Scientific Research contract No. F49620-84-C-0100. The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research of the U.S. Government.

Final Report
Knowledge Delivery Research

William C. Mann
USC Information Sciences Institute

October, 1986

Abstract

The goal of knowledge delivery research is to create a technology of authorship by computer. Existing technology is all in the laboratory stage, and is limited to very small, rigidly constrained texts. This research project has focused on two kinds of developments intended to overcome these limits: 1) expanding the notation and practices of knowledge representation so that a wider range of knowledge can be rendered in natural language, and 2) creating a theory of text structure that is suitable as a basis for writing programs that design texts.

This is the final research report for AFOSR contract FQ8671-84-01007, part of an ongoing research program at USC Information Sciences Institute. The research is not complete, and is being continued under contract F49620-87-C-0005. This report represents the research accomplishments in the interval of August 15, 1984 to August 14, 1986.

1 Goals

The general goal of Knowledge Delivery Research is to create a technology of authorship by computer, so that computers can represent their knowledge freely in written English. Since existing technology is limited to small texts, a major subgoal is to create reusable, size-insensitive methods for programs to use in creating texts. These goals are being pursued using the methods of Artificial Intelligence, with heavy input from Linguistics.

2 Accomplishments

2.1 Knowledge Notation

The English language, and much of human knowledge, is organized around concepts of actions and their participants. Terms such as "creator", "recipient" and "owner" are not really definable apart from some sort of action orientation. In contrast, the knowledge notations of AI and mathematics do not give any special organizational place to actions or their participants, and there are no strong precedents for representing such knowledge.

This creates several difficulties for expert systems and for English language knowledge delivery:

1. Information about actions tends to get represented inconsistently, with the result that uniform information processing methods yield inconsistent results.

2. Systematic expression in English of knowledge about actions is made difficult.

3. Computer processes which interpret action-oriented English tend to lose information.

In regularizing knowledge, the uniformity can come either from a restrictive notation or from uniform conventions of use of a less restrictive notation. The general knowledge notations of AI and mathematics are not easily converted into restrictive notations that enforce uniform or canonical representation of actions. To do so would dilute some of their carefully developed advantages. We are therefore developing conventions of use of existing notations for representing knowledge of actions.

One of the basic representational strategies of AI is to use a taxonomy of concept types. Commonly these are organized around the entities of the domain of processing, such as chemical reactions or diagnosis of electronic equipment. Generic concepts, such as action, are seldom used. Such concepts are typically not in conflict with the domain, but in early design they are often considered superfluous.

We are building a highly abstract taxonomy, including action concepts; it is intended as a base upon which more detailed (domain) concepts can be added. This taxonomy, called Upper Structure, provides for a wide range of the organizing concepts of English, not just action and participants. It is in NIKL, a well known AI knowledge notation [Schmolze & Brachman 82]. In NIKL, concepts have not only positions in a hierarchy, but also roles which related them to other concepts. NIKL provides inheritance of roles in the hierarchy. Figure 1 shows the Upper Structure taxonomy of concepts, without roles, and Figure 2 shows action with related concepts.

Finding an appropriate hierarchic representation is made difficult both by the interactions of representational choices and by the high level of generality desired. The version shown has been used in an English generation program, Penman, which has been tested with several knowledge domains, including computer mail and calendars, assistance to Penman users, and a weapons data base.

2.2 Text Organization

Much of the knowledge in modern computer programs is too complex to express in single sentences, making it necessary to organize larger texts for knowledge delivery. Unfortunately there is no well established technology of text organization at a level of detail that could guide computer programming.

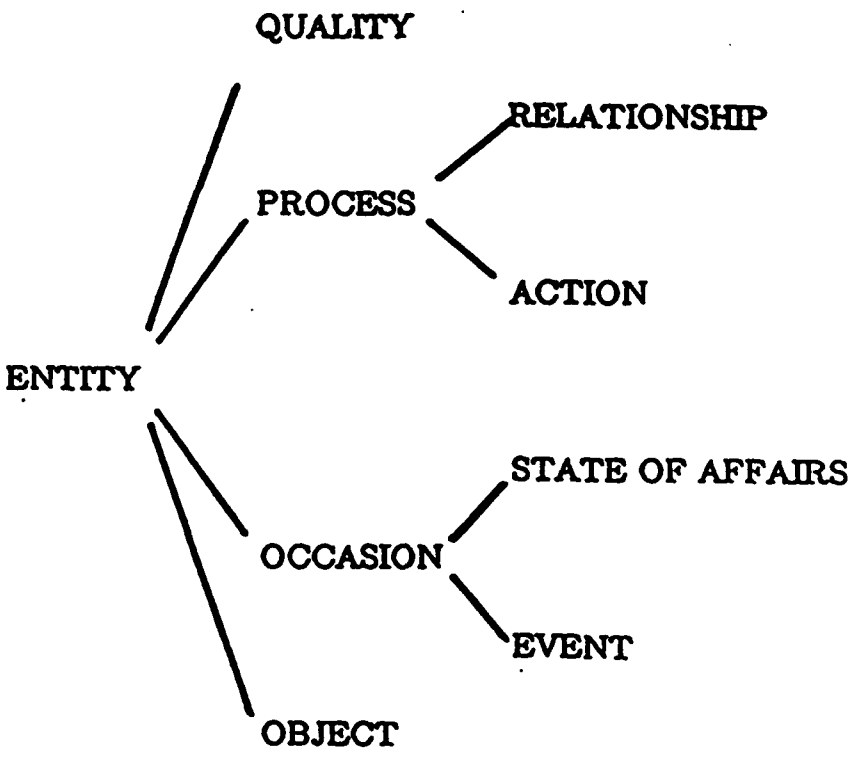


Figure 1: Hierarchy of Upper Structure Categories

We are therefore working to create a basic theoretical knowledge of text organization. We have named one major portion of this Rhetorical Structure Theory (RST). RST now provides a computationally useful account for

1. What the functional parts of a text are,
2. How the parts are related,
3. What the predictable consequences of putting parts adjacent are,
4. How the order of parts is determined,
5. Why the writer includes particular parts in the text.

Recent work has put RST on a new definitional foundation [Mann & Thompson 87]. The benefits include these:

1. The relationships between semantic and communication constraints on texts are identified, with the two types coresident in definitions in same terminology; previous methods left them unrelated.

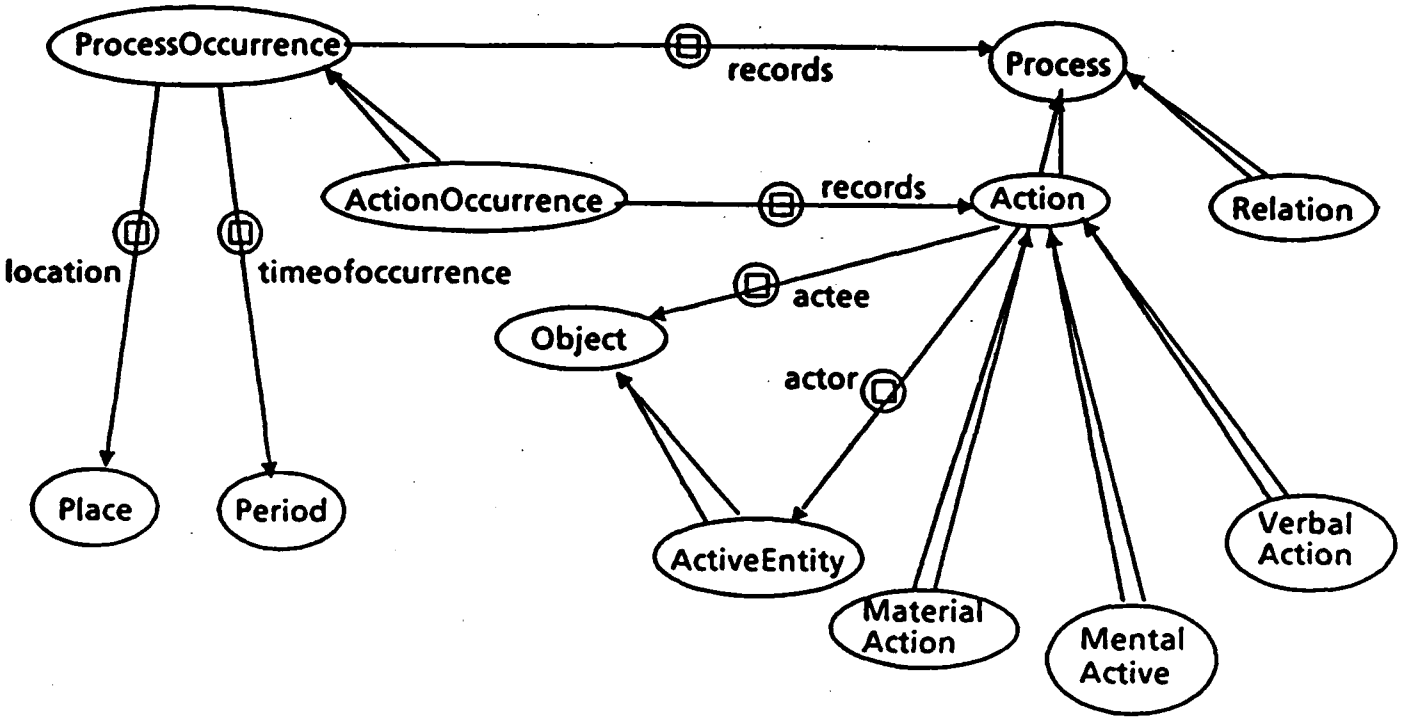


Figure 2: Action and Related Concepts

2. Direct clausal assertion, as in "John broke the bottle", is related to implicit assertion, as in the pair "John dropped the bottle. It broke." , which asserts cause. Previous frameworks left them unrelated.

3. Enumerates varieties of implicit assertions in texts, potentially leading to program control of implicit assertion in multisentence text.

4. Identifies the theoretical source of implicit assertions, and shows how they can be derived from related knowledge. (In our previous work they were arbitrary correlates of text structure.)

5. Identifies a limited set of points of subjectivity in text analysis, and shows that the corresponding subjectivity is not necessary in text synthesis.

Recent work has created a RST Structure Construction Procedure, a procedural descriptive framework which can be refined into programmable text structure planners for particular knowledge domains and tasks [Ford & Mann 86a], [Ford & Mann 86b]. This construction procedure is step 2 in the larger picture presented in Figure 3.

There is a companion plan, a utilization procedure which shows how the planner can be applied to particular domains. As part of developing text construction techniques, this utilization procedure is currently undergoing its first extensive test.

3 Remaining Goals

Corresponding to the two areas of accomplishment described above, there are two areas of continuing need in text generation research: knowledge representation and text design.

In the knowledge representation area we have identified a collection of about 10 critical limits on present representational notations and techniques. Of these we are actively addressing time, actions and participants and propositional relations, which are all crucial both because they are prominent in English syntax (as tense, transitivity and conjunctive relations) and also because they are prominent in representing the sorts of computer operations that arise in AI expert systems.

Our approach to these involves a combination of development of effective notational conventions and development of the notations themselves. Experiments are under way, but not yet ready for evaluation. We plan to address others when these three are relatively well developed.

In the text design area, a good start has been made in automatic design of text structure, through constructive RST. Part of the progress has come in factoring the text design problem into parts, and separating structure building processes from processes that work with other characteristics of the text being designed. Figure 3 shows our current factoring.

The processes other than structure building are in a much more rudimentary state of development. For most of them, trivialized approaches exist which can be used to generate basic texts, but the issues must be investigated in order to develop high fluency in generation.

4 Technical Summary

Substantial progress has been made in creating a technology with which computers can represent their knowledge freely in written English. The methods are being made reusable and compatible with English partly by encoding generic knowledge in an AI-style Upper Structure.

Methods for designing texts are being developed based on Rhetorical Structure Theory, which provides extensive information about the available alternatives of text organization and the effects of using them.

Experiments on the representation of time, actions and participants and propositional relations have been begun but not completed. Similarly, experiments in automatic structure building have been begun. These need to be carried into the evaluation and process improvement stages. Additional work is needed in several other quality-limiting areas of knowledge representation which have already been identified, and in non-structural aspects of automatic text design.

1. Before building RST structure:

   a. General decisions about what to accomplish, what knowledge to use. This yields a body of material to convey.

   b. Identification of the audience.

2. RST structure building, including

   a. Organizing the given body of material.

   b. Supplementing it as needed, with evidence, concessives, circumstantials, antithesis, contrast and other supporting material.

   c. Ordering nuclei and satellites.

3. After building RST structure:

   a. Theme control,

   b. Sentence scope,

   c. Conjunction uses,

   d. Lexical choice,

   e. Formulaic text, e.g. "Sincerely yours,",

   f. Grammatical realization.

Figure 3: RST Structure Construction and Related Processes

5 Publications

The publications for the contract period, together with the submitted publications and reports which are currently beyond first draft form, and the major public presentations of the project work, are as follows:

1. [Mann & Thompson 85] -- "Assertions from Discourse Structure." It turns out that RST relations have assertional properties, in which they convey information from the discourse structure, distinct from the clausal assertions of the text. In [Mann & Thompson 86] this phenomenon is identified, and in [Mann & Thompson 85] the link between RST and the phenomenon is made.

2. [Mann 86a] -- The 1986 development status of RST is described in this project report, which will be part of a 1987 book based on the Third International Workshop on Text Generation [Kempen 86].

3. [Thompson & Mann 87] -- "Clause Combining -- Antithesis Relations at Large Scale and Clause Scale." In this paper, clausal-level and large-scale antithesis relations are found to rest on the same discourse configurations. Significant in unifying the planning methods of large and small scales. Accepted for inclusion in an edited book.

4. [Matthiessen & Thompson 86] -- "Clause Combining -- Hypotaxis, Embedding and "Subordination". In this paper, RST is used to establish that so-called subordination is a composite category consisting of two different phenomena. A project report not yet printed, also accepted for publication in an edited book in 1987.

5. [Matthiessen 87] -- The limitations of systemic notation and the phenomena which are thereby made hard to represent are explored. A conference paper already included in an announced book, forthcoming as a project report.

6. [Kasper 87] -- The benefits of representing systemic grammars in functional unification form are explored, along with problems and solutions. A conference paper already included in an announced book, forthcoming as a project report.

7. [Cumming 86] -- The lexicon is a crucial and distictive part of generation technology. This paper compares the lexical technologies of a wide variety of existing generation systems. To appear in an edited book based on the workshop. Currently being printed as a project report.

8. [Sondheimer 86] -- This paper describes methods for linking the Nigel grammar to the Upper Structure abstractions hierarchy, using a first-order predicate calculus input language for the text generator, and the KL-TWO knowledge notation as an intermediary.

9. [Mann 86b] -- This paper compares the text structuring methods of RST and McKeown's TEXT system. It was presented as an invited paper at COLING86, and has been accepted to appear in an edited book in 1987.

10. [Mann & Thompson 87] -- This large paper features the new definitional basis of RST. It is scheduled to appear in an edited book.

11. [Ford & Mann 86a] -- This paper, related to the companion paper below, describes the current design of a process of text structure creation using RST. To appear as a project report.

12. [Ford & Mann 86b] -- The process of text structure creation has been tested in a variety of ways. The tests, together with some refinements which they suggested, are described in this paper. To appear as a project report.

13. [Poulton 86] -- Nigel is a large systemic grammar of English represented in a computer program for sentence generation. This report is documentation for the program user. To appear as a project report.

14. [Matthiessen 86] -- Nigel generates sentences relative to an environment of knowledge of the world and the text plan. This paper describes the organization of that environment. To appear as a project report.

6 Outside Research Based on the Knowledge Delivery Project

We are aware of the following outside research based on this research project which has been published or begun:

1. [Fox 84] -- "Predicting Anaphora." In this completed Phd dissertation, RST is used in predicting when anaphora will be used.

2. [Noel 86] -- The methods of RST analysis have been combined with complementary methods in a linguistic characterization of the presentational methods of the BBC World News Service broadcasts. A completed and published dissertation, not PhD.

3. [Cui 85] Contrastive Rhetoric -- Comparing Structures of Essays in Chinese and English. This is a completed MA dissertation, based on RST.

4. Clause Combining -- Switch Reference in Quechua. This is a PhD dissertation in progress at UCLA, based on RST.

In addition to these identified presentations, there were presentations by Dr. Mann and Dr. Sondheimer at the most recent AAAI, ACL, COLING, FJCC, German AI Conference and (January 1987) TINLAP conferences which represented this research in substantial ways but were not confined to it.

7 Project Personnel

The technical staff members of this project were:[1]

William C. Mann — Principal Investigator

---

[1]This project has benefitted by interaction with another ongoing project on text generation sponsored by DARPA under contract MDA903 81 C 0335. The list of project personnel is unusually long because they have all participated in both efforts.

Norman K. Sondheimer	Co-project-leader
Robert Albano	
Susanna Cumming	
Cecilia Ford	Thomas Galloway
Robert Kasper	Christian M. I. M. Matthiessen
Shari Naberschnig	Lynn Poulton
Sandra A. Thompson	George Vamos
Richard Whitney	

Support staff were:

Pam Andro	Heidi Julian
Lisa Trentham	

References

[Cui 85] Cui, Songren, *Comparing Structures of Essays in Chinese and English*, Ph.D. thesis, UCLA, 1985.

[Cumming 86] Cumming, Susanna, *The Lexicon in Text Generation*, USC/Information Sciences Institute, Technical Report ISI/RR-86-168, 1986.

[Ford & Mann 86a] Ford, Cecilia and William C. Mann , *Process Design for a Rhetorical Structure Planner*, USC/Information Sciences Institute, Technical Report , 1986.

[Ford & Mann 86b] Ford, Cecilia and William C. Mann, *Testing a Rhetorical Structure Planner*, USC/Information Sciences Institute, Technical Report , 1986.

[Fox 84] Fox, Barbara A., *Discourse Structure and Anaphora in Written and Conversational English*, Ph.D. thesis, UCLA, 1984.

[Kasper 87] Kasper, Robert , "Systemic Grammar and Functional Unification Grammar," in Benson, James D. and William S. Greaves (ed.), *Functional Perspectives on Discourse: Selected Papers from the 12th International Systemic Workshop*, Ablex, Norwood, New Jersey, 1987.

[Kempen 86] Kempen, Gerard, (ed.), *Proceedings of the Third International Workshop on Text Generation*, , Nijmegen, The Netherlands, 1986.

[Mann 86a] Mann, William C., "Rhetorical Structure Theory: Description and Construction of Text Structures," in Gerard Kempen (ed.), *Proceedings of the Third International Workshop on Text Generation*, Nijmegen, The Netherlands, August 1986.

[Mann 86b] Mann, William C., "Text Generation: The Problem of Text Structure," , August 1986. The contents of this COLING '86 lecture have been accepted for inclusion in an edited book on text generation, and will also appear as an ISI report.

[Mann & Thompson 85] Mann, William C. and Sandra A. Thompson, "Assertions from Discourse Structure," in *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*, Berkeley Linguistic Society, Berkeley, 1985. Also available as ISI/RS-85-155.

[Mann & Thompson 86] Mann, William C. and Sandra A. Thompson, "Relational Propositions in Discourse," *Discourse Processes* 9, (1), January-March 1986, 57-90. Also available as ISI/RR-83-115.

[Mann & Thompson 87] Mann, William C. and Thompson, Sandra A., "Rhetorical Structure Theory: A Theory of Text Organization," in Livia Polanyi (ed.), *Discourse Structure*, Ablex, Norwood, N.J., 1987. To Appear.

[Matthiessen 86] Matthiessen, Christian M. I. M., *The Organization of the Environment of a Text Generation Grammar*, USC/Information Sciences Institute, Technical Report ISI/RR-86-177, 1986.

[Matthiessen 87] Matthiessen, Christian M. I. M., "Representational Issues in Systemic Functional Grammar," in Benson, James D. and William S. Greaves (ed.), *Functional Perspectives on Discourse: Selected Papers from the 12th International Systemic Workshop*, Ablex, Norwood, New Jersey, 1987.

[Matthiessen & Thompson 86] Matthiessen, Christian and Thompson, Sandra A., "The Structure of Discourse and "Subordination"," in Haiman and Thompson (eds.), *Clause Combining in Grammar and Discourse*, Benjamins, Amsterdam, 1986. To Appear

[Noel 86] Noel, Dirk, *Towards a Functional Characterization of the News of the BBC World News Service.* Antwerp, Belgium, 1986. Antwerp Papers in Linguistics, Number 49.

[Poulton 86] Poulton, Lynn, *The Nigel English Sentence Generator: Users Manual*, USC/Information Sciences Institute, Technical Report , 1986.

[Schmolze & Brachman 82] Schmolze, James G., and Ronald J. Brachman, *Proceedings of the 1981 KL-ONE Workshop*, Fairchild Corporation, Technical Report 618, May 1982.

[Sondheimer & Nebel 86] Sondheimer, Norman K. and Bernhard Nebel, "A Logical-Form and Knowledge-Base Design for Natural Language Generation," in *AAAI-86, Proceedings of the National Conference on Artificial Intelligence*, AAAI, Philadelphia, PA, August 1986.

[Thompson & Mann 87] Thompson, Sandra A. and William C. Mann, "Antithesis: A Study in Clause Combining and Discourse Structure ," in *Nominal Book Title*, Nominal Press, 1987. Submitted for inclusion in a festschrift book. Publication to be announced.

END

1-87

DTIC